Homework 1

Dennis Wang
CSCI 1470 Deep Learning

September 12, 2025

1 A. Matrix Multiplication

1.1 Task 1

1. Prove that (2) + (3) implies (4)

   Given $M_1 \in \mathbb{R}^{d \times c}$ and suppose vector $x \in \mathbb{R}^{c \times 1}$

   Then by (3), $Mx = M \begin{bmatrix} x_0 \\ \vdots \\ x_{c-1} \end{bmatrix} = \begin{bmatrix} M_0 x \\ \vdots \\ M_{r-1} x \end{bmatrix}$ and each row $M_i x$ is a real scalar (2).

   Therefore $\begin{bmatrix} M_0 x \\ \vdots \\ M_{r-1} x \end{bmatrix} \in \mathbb{R}^{r \times 1}$

2. Prove that (4) implies (5)

   Given $M_1 \in \mathbb{R}^{d \times c}$ and $M_2 \in \mathbb{R}^{r \times d}$

   Then $M_2 M_1 \in \mathbb{R}^{r \times c}$

   And suppose vector $x \in \mathbb{R}^{c \times 1}$, then
   $M_2 M_1 x \in \mathbb{R}^{r \times 1}$ or $\mathbb{R}^r$

2 B. Differentiation

2.1 Task 2

1. Solve $\frac{\partial}{\partial y} ln(x^5/y^2)$

   $\frac{\partial}{\partial y} ln(x^5/y^2) = \frac{\partial}{\partial y} ln(x^5 y^{-2}) = \frac{1}{x^5 y^{-2}} \cdot \frac{\partial}{\partial y} x^5 y^{-2} = \frac{-2x^5 y^{-3}}{x^5 y^{-2}} = \frac{-2}{y}$

2. Solve for a valid j and all valid i:

$$\frac{\partial}{\partial x_j} ln \left[\sum_i x_i y_i \right]$$

Let $g_1(x) = \sum_i x_i y$
$\frac{\partial}{\partial x_j} ln[g_1(x)]$
$= \frac{1}{g_1(x)} \cdot \frac{\partial}{\partial x_j} g(x)$
$= \frac{1}{\sum_i x_i y} \cdot \sum_i y_i$
or it all equals 0 because we're trying to find the derivative w.r.t x_j which isn't in the equation?

3 C. Jacobians

3.1 Task 3: Higher-Dimensional Jacobians and Multi-Stage Function Analysis

3.1.1 Part A: Jacobian Computation and Analysis

1. $J_g = \begin{bmatrix} \frac{\partial g_1}{\partial u} & \frac{\partial g_1}{\partial v} & \frac{\partial g_1}{\partial w} \\ \frac{\partial g_2}{\partial u} & \frac{\partial g_2}{\partial v} & \frac{\partial g_2}{\partial w} \\ \frac{\partial g_3}{\partial u} & \frac{\partial g_3}{\partial v} & \frac{\partial g_3}{\partial w} \end{bmatrix} = \begin{bmatrix} e^u & \frac{e^v}{1+e^v} & 0 \\ 2uw & 0 & u^2+1 \\ 0 & \frac{1}{e^v(1+e^{-v})^2} & 3w^2 \end{bmatrix}$

2. $J_g(1,0,1) = \begin{bmatrix} e & 1 & 0 \\ 2 & 0 & 2 \\ 0 & \frac{1}{4} & 3 \end{bmatrix}$

3.1.2 Part B: Function Composition and Chain Rule

1. $p(g(f(s,t)))$ is not valid because function g outputs 3 values but function p only takes 2 values for input.

2. $g(f(s,t))$

   i $g(s^2t, s+e^t, ln(1+s^2)) = \begin{bmatrix} e^{s^2t} + ln(1+e^{s+e^t}) \\ ln(1+s^2)(s^4t^2+1) \\ \frac{1}{1+e^{-s-e^t}+ln(1+s^2)^3} \end{bmatrix}$

   ii $\frac{\partial}{\partial(s,t)} g(f(s,t)) = \begin{bmatrix} 2ste^{s^2t} + \frac{e^{s+e^t}}{1+e^{s+e^t}} & s^2e^{s^2t} + \frac{e^{t+s+e^t}}{1+e^{s+e^t}} \\ 4s^3t^2ln(1+s^2) + \frac{2s^5t^2+2s}{1+s^2} & 2ts^4ln(1+s^2) \\ \frac{e^{-s-e^t}}{(1+e^{-s-e^t})^2} + \frac{6sln(1+s^2)^2}{1+s^2} & \frac{e^{-s-e^t+t}}{(1+e^{-s-e^t})^2} \end{bmatrix}$

   iii $J_g(f(s,t)) \cdot J_f(s,t) =?$

Let $u = s^2t, v = s + e^t, w = ln(1 + s^2)$

$$J_g = \begin{bmatrix} e^u & \frac{e^v}{1+e^v} & 0 \\ 2uw & 0 & u^2+1 \\ 0 & \frac{1}{e^v(1+e^{-v})^2} & 3w^2 \end{bmatrix} = \begin{bmatrix} e^{s^2t} & \frac{e^{s+e^t}}{1+e^{s+e^t}} & 0 \\ 2(s^2t)ln(1+s^2) & 0 & s^4t^2+1 \\ 0 & \frac{1}{e^{s+e^t}(1+e^{-s-e^t})^2} & 3ln(1+s^2)^2 \end{bmatrix}$$

$$J_f(s,t) = \begin{bmatrix} \frac{\partial f_1}{\partial s} & \frac{\partial f_1}{\partial t} \\ \frac{\partial f_2}{\partial s} & \frac{\partial f_2}{\partial t} \\ \frac{\partial f_3}{\partial s} & \frac{\partial f_3}{\partial t} \end{bmatrix} = \begin{bmatrix} 2st & s^2 \\ 1 & e^t \\ \frac{2s}{1+s^2} & 0 \end{bmatrix}$$

$$J_g(f(s,t)){\cdot}J_f(s,t) = \begin{bmatrix} e^{s^2t} & \frac{e^{s+e^t}}{1+e^{s+e^t}} & 0 \\ 2(s^2t)ln(1+s^2) & 0 & s^4t^2+1 \\ 0 & \frac{1}{e^{s+e^t}(1+e^{-s-e^t})^2} & 3ln(1+s^2)^2 \end{bmatrix} \cdot \begin{bmatrix} 2st & s^2 \\ 1 & e^t \\ \frac{2s}{1+s^2} & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 2ste^{s^2t} + \frac{e^{s+e^t}}{1+e^{s+e^t}} & s^2e^{s^2t} + \frac{e^{s+t+e^t}}{1+e^{s+e^t}} \\ 4s^3t^2ln(1+s^2) + \frac{2s^5t^2+2s}{1+s^2} & 2s^4tln(1+s^2) \\ \frac{1}{e^{s^s+e^t}(1+e^{-s-e^t})^2} + \frac{6sln(1+s^2)^2}{1+s^2} & \frac{e^t}{e^{s+e^t}(1+e^{-s-e^t})^2} \end{bmatrix}$$

equivalent!

3.2 Task 4: Element-wise Functions

1. (a) $r'(x) = 0ifx \leq 0$
   $r'(x) = 1ifx > 0$

   (b) $J_r = \begin{bmatrix} r'(x_1) & 0 & 0 \\ 0 & r'(x_2) & 0 \\ 0 & 0 & r'(x_3) \end{bmatrix}$

2. $J_A = \begin{bmatrix} f'_A(x_1) & 0 \\ 0 & f'_A(x_2) \end{bmatrix} = \begin{bmatrix} 2x_1 & 0 \\ 0 & 2x_2 \end{bmatrix}$

   $J_B = \begin{bmatrix} 2x_1 & 1 \\ 1 & 2x_2 \end{bmatrix}$ J_A is diagonal because the outputs of function A only depend on their corresponding input

4 D. Probability

4.1 Task 5

i 0.5

ii No because assigning 0.5 probability to each prediction makes the classifier effectively useless for any real world application even if prediction probability matches the distribution of the cats and dogs in the dataset. It would be akin to randomly guessing each prediction, which would lead to really poor performance metrics.

4.2 Task 6: Essential Probability Calculations

4.2.1 Part A: Basic Computations

1. $\mathbb{E}[Z] = (-2 \times 0.3) + (0 \times 0.4) + (2 \times 0.3) = 0$
   $\mathbb{V}[Z] = (-2-0)^2(0.3) + (0) + (2-0)^2(0.3) = 2.4$
   $\mathbb{E}[Z^2] = \mathbb{V}[Z] + \mathbb{E}[Z]^2 = 2.4$

2. Given $X \sim N(2, 9), Y \sim N(-1, 4)$ and are independent:

   (a) $X + Y \sim N(1, 13)$

   (b) $3X - 2Y + 5 \sim N(13, 97)$
   $\mathbb{V}[3X] = 9\mathbb{V}[X] = 9 \times 9 = 81$
   $\mathbb{V}[2Y] = 4\mathbb{V}[Y] = 4 \times 4 = 16$
   $81 + 16 = 97$

4.2.2 Part B: Matrix-Vector Products with Random Matrices

3. (a) $\mathbb{E}[Av] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} A_{11} - A_{12} \\ A_{21} - A_{22} \\ A_{31} - A_{32} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

   (b) $\mathbb{V}[(Av)_1] = \mathbb{V}[(A_{11} \cdot 1 + A_{12} \cdot (-1))] = \mathbb{V}[A_{11}] + \mathbb{V}[-A_{12}] = 1 + 1 = 2$

   (c) If v is any vector with $||v||^2 = c$, then $\mathbb{V}[(Av)_i] = v^2\mathbb{V}[A_i] = v^2 \cdot 1 = c$ for any component i.

4.2.3 Part C: Optimization from Probabilistic Assumptions

4. (a) probability density for y= $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$
   Given $y_i \sim N(2x_i + 3, 1)$, let $\mu_1 = 2x_i + 3 = 2(1) + 3 = 5, \sigma^2 = 1, y_i = 4.8$
   $\frac{1}{\sqrt{2\pi \cdot 1}}e^{-\frac{(4.8-5)^2}{2.1}} = \frac{1}{2\pi}e^{-0.02}$

   (b) $y_i - ax_i - b$ is shifting the entire y=ax+b equation to one side of the equal sign, so the expected value is 0. However, if we include the error term $\epsilon_i \sim N(0, 1)$, then $y_i - ax_i - b$ would equal the error term. We want to minimize the squared sum of the error terms because it should be close to the expected value of 0.

4.2.4 Part D: Averaging Independent Quantities

5. (a) $\mathbb{E}[\overline{M}] = \mathbb{E}[\frac{1}{16}\sum_{i=1}^{16} M_i] = [\frac{1}{16}\mathbb{E}\sum_{i=1}^{16} M_i] = \frac{1}{16} \cdot 16 \cdot \mu = \mu$

   (b) Given $\mathbb{V}[M_i] = \sigma^2$, then
   $\mathbb{V}[\overline{M}] = \mathbb{V}[\frac{1}{16}\sum_{i=1}^{16} M_i] = [\frac{1}{16^2}\sum_{i=1}^{16} \mathbb{V}M_i] = \frac{1}{16^2} \cdot 16 \cdot \sigma^2 = \frac{\sigma^2}{16}$

   (c) $\mathbb{V}[\overline{M}] = \frac{1}{4} \cdot \frac{\sigma^2}{16} = \frac{\sigma^2}{64}$